Homology modelling of the Lactococcus lactis leader peptidase NisP and its interaction with the precursor of the lantibiotic nisin

Siezen, Roland J.; Rollema, Harry S.; Kuipers, Oscar P.; Vos, Willem M. de

Published in:
%22Protein Engineering%2C Design and Selection%22

DOI:
10.1093/protein/8.2.117

Protein Engineering vol.8 no.2 pp.117–125, 1995

Homology modelling of the *Lactococcus lactis* leader peptidase NisP and its interaction with the precursor of the lantibiotic nisin

Roland J.Siezen[1], Harry S.Rollema, Oscar P.Kuipers and Willem M.de Vos

Department of Biophysical Chemistry, NIZO, PO Box 20 Ede, 6710BA, The Netherlands

[1]To whom correspondence should be addressed

A model is presented for the 3-D structure of the catalytic domain of the putative leader peptidase NisP of *Lactococcus lactis*, and the interaction with its specific substrate, the precursor of the lantibiotic nisin. This homology model is based on the crystal structures of subtilisin BPN′ and thermitase in complex with the inhibitor eglin. Predictions are made of the general protein fold, inserted loops, Ca^{2+} binding sites, aromatic interactions and electrostatic interactions of NisP. Cleavage of the leader peptide from precursor nisin by NisP is the last step in maturation of nisin. A detailed prediction of the substrate binding site attempts to explain the basis of specificity of NisP for precursor nisin. Specific acidic residues in the S1 subsite of the substrate binding region of NisP appear to be of particular importance for electrostatic interaction with the P1 Arg residue of precursor nisin after which cleavage occurs. The hydrophobic S4 subsite of NisP may also contribute to substrate binding as it does in subtilisins. Predictions of enzyme–substrate interaction were tested by protein engineering of precursor nisin and determining susceptibility of mutant precursors to cleavage by NisP. An unusual property of NisP predicted from this catalytic domain model is a surface patch near the substrate binding region which is extremely rich in aromatic residues. It may be involved in binding to the cell membrane or to hydrophobic membrane proteins, or it may serve as the recognition and binding region for the modified, hydrophobic C-terminal segment of precursor nisin. Similar predictions for the tertiary structure and substrate binding are made for the highly homologous protein EpiP, the putative leader peptidase for the lantibiotic epidermin from *Staphylococcus epidermidis*, but EpiP lacks the aromatic patch. Based on these models, protein engineering can be employed not only to test the predicted enzyme–substrate interactions, but also to design lantibiotic leader peptidases with a desired specificity.

Key words: leader peptidase/molecular modelling/nisin/serine protease/substrate specificity/subtilase

Introduction

Nisin is an antibacterial peptide, belonging to the class of lantibiotics, that is produced and secreted by several strains of *Lactococcus lactis*. It strongly inhibits the growth of a broad range of Gram-positive bacteria (Hurst, 1981; Delves-Broughton, 1990). Nisin is ribosomally synthesized as a 57 residue peptide (pre-nisin) which first undergoes a number of modifications, generating precursor nisin (Figure 1) that is subsequently secreted and processed before it becomes active. The highly modified mature peptide of 34 amino acid residues contains several uncommon features, such as lanthionine and β-methyllanthionine residues forming five intramolecular thioether bridges, and the dehydrated residues dehydro-alanine and dehydro-butyrine, derived from Ser and Thr residues, respectively (Gross and Morell, 1971; Hurst, 1981). It has been shown that removal of the highly charged leader peptide of 23 residues is the last step in the maturation of nisin, and that this step requires a specific, extracellular leader peptidase (Van der Meer *et al.*, 1993, 1994).

The genes required for biosynthesis of nisin in *L.lactis* NIZO R5 are organized in a gene cluster, located on the nisin–sucrose transposon Tn*5276* (Rauch and De Vos, 1992; Kuipers *et al.*, 1993a; Van der Meer *et al.*, 1993). Sequence analysis shows that one of these genes (*nisP*) encodes a protein of 682 amino acid residues with characteristics of a pre-pro-protein designated to become a secreted, membrane-bound serine protease (Figure 2). This putative leader peptidase NisP has an N-terminal pre-pro sequence of ~200 residues, a central segment of ~350 residues which shows significant homology to members of the family of subtilisin-like serine proteases, a spacer of ~80 residues and finally a C-terminal consensus membrane anchor sequence (Van der Meer *et al.*, 1993). The general-purpose, cell-envelope proteinase of *L.lactis* has similar structural characteristics but is much larger (Vos *et al.*, 1989). The catalytic domain of NisP is most similar (42% identical residues) to the putative protease EpiP encoded in the gene cluster for biosynthesis of epidermin, a related lantibiotic from *Staphylococcus epidermidis* (Schnell *et al.*, 1992).

Gene deletion studies have demonstrated that in the absence of the *nisP* gene a fully modified, inactive precursor of nisin, to which the leader peptide is still attached, is efficiently secreted by *L.lactis*. This precursor can be processed to active nisin *in vitro* when either *nisP*-expressing lactococcal cells or trypsin is added, both leading to cleavage after the Arg(–1) residue of the leader peptide (Van der Meer *et al.*, 1993). Heterologous expression of the *nisP* gene in *Escherichia coli* produced an active protease that correctly cleaved off the nisin leader peptide to generate active nisin *in vitro* (Van der Meer *et al.*, 1993); the size of the active protease NisP produced by *E.coli* was ~54 kDa, which corresponds well to that of the mature form expected in *L.lactis*. These studies clearly demonstrated that NisP is the protease required for removal of the nisin leader peptide, and that NisP is located extracellularly in *L.lactis*.

Over 100 members of the family of subtilisin-like serine proteinases or subtilases are presently known (Siezen *et al.*, 1991; R.J.Siezen, manuscript in preparation), but only a limited number of 3-D structures has been determined, e.g. subtilisin BPN′ (Hirono *et al.*, 1984; McPhalen *et al.*, 1985; McPhalen and James, 1988; Heinz *et al.*, 1991; Takeuchi *et al.*, 1991a,b), subtilisin Carlsberg (Bode *et al.*, 1986; McPhalen and James, 1988), subtilisin BL (Godette *et al.*, 1992), high-alkaline protease (Betzel *et al.*, 1992; Sobek *et al.*, 1992; Van der Laan *et al.*, 1992; Lange *et al.*, 1994), mesentericopeptidase (Dauter

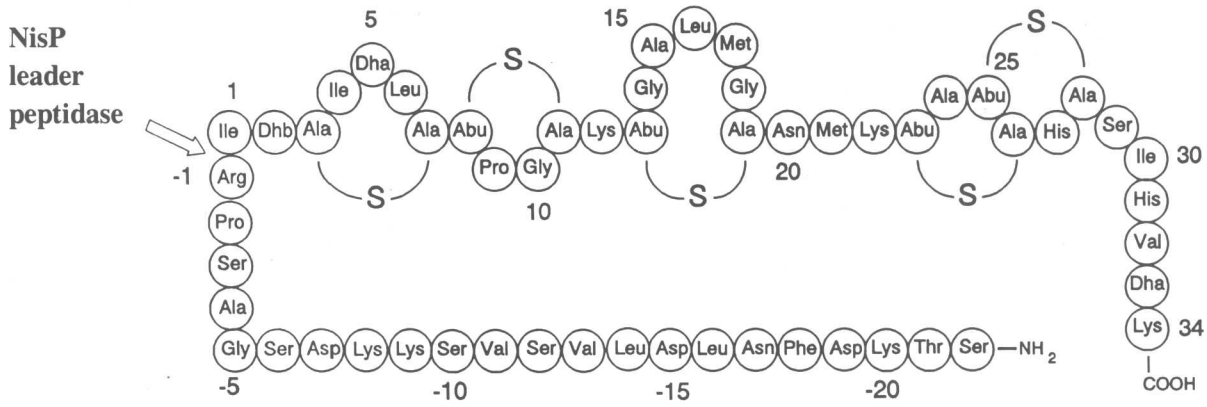

Fig. 1. Schematic representation of precursor nisin A (the nisin Z variant has a His27Asn substitution). The cleavage site by NisP is indicated by an arrow. The pro-nisin part is fully modified, as indicated by the modified residues. Dha, dehydroalanine; Dhb, dehydrobutyrine; Ala-S-Ala, lanthionine; Abu-S-Ala, β-methyllanthionine.

et al., 1991), thermitase (Gros *et al.*, 1989; Betzel *et al.*, 1990; Teplyakov *et al.*, 1990) and proteinase K (Betzel *et al.*, 1988a, 1990). Residues in subtilases that are in contact with substrate or inhibitor have been identified from crystal structures and modelling of enzyme–inhibitor complexes (Hirono *et al.*, 1984; McPhalen *et al.*, 1985; Bode *et al.*, 1987; Betzel *et al.*, 1988b, 1993; McPhalen and James, 1988; Gros *et al.*, 1989; Dauter *et al.*, 1991; Heinz *et al.*, 1991; Takeuchi *et al.*, 1991a,b) and enzyme–substrate complexes (Wells *et al.*, 1987a), and from protein engineering studies of subtilisin BPN′ (Estell *et al.*, 1986; Russell and Fersht, 1987; Russell *et al.*, 1987; Wells *et al.*, 1987a,b). In each case the mode of substrate–inhibitor binding appears to be essentially the same, and this provides a sound basis for predicting enzyme–substrate interactions in other members of this family. Homology modelling has been used to make predictions for the catalytic domain of other subtilases concerning structure, stability, substrate specificity, etc. (Siezen *et al.*, 1991, 1993, 1994).

Here we describe the homology modelling of the catalytic domain of NisP in complex with precursor nisin, based on sequence alignment and the known X-ray structures of subtilisin BPN′ and thermitase in complex with the inhibitor eglin.

Materials and methods

Coordinates

The coordinates of the crystal structures of subtilisin BPN′ in complex with the inhibitor Arg45-eglin (Heinz *et al.*, 1991), and thermitase in complex with eglin (Gros *et al.*, 1989; Brookhaven Protein Data Bank code 1TEC), were used; these structures were chosen because thermitase and subtilisin have the highest sequence identity with NisP, and the Arg45-eglin is suitable for modelling the substrate precursor nisin which also has an Arg(P1) residue (see below). Four calcium binding sites are known in the subtilase family: a strong (Ca1), a medium-strength (Ca2) and two weaker binding sites (Ca3, Ca4; Betzel *et al.*, 1988a; Gros *et al.*, 1989; Tepylakov *et al.*, 1990; Siezen *et al.*, 1991). The Ca1 and Ca2 ions are present in the thermitase structure used, whereas subtilisin BPN′ has the Ca1 and Ca3 ions and proteinase K has the Ca3 and Ca4 ions.

The coordinates of nisin in aqueous solution were determined by high-resolution ^{1}H-NMR (Van de Ven *et al.*, 1991).

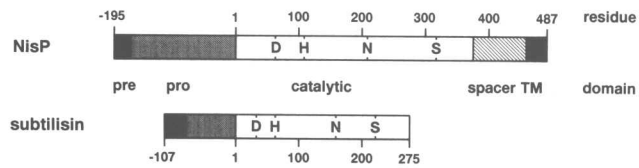

Fig. 2. Schematic representation of precursor NisP and comparison with precursor subtilisin. Segments or domains with different functions are shaded. TM, transmembrane anchor. The relative positions of active-site and oxy-anion hole residues in the catalytic domain are indicated by D (Asp), H (His), N (Asn) and S (Ser). The cleavage site within the pre-pro-peptide of NisP is putative (Van der Meer *et al.*, 1993).

Sequence alignment

The amino acid sequences of the catalytic domain of *L.lactis* NisP (Van der Meer *et al.*, 1993) and *S.epidermidis* EpiP (Schnell *et al.*, 1992) were included in the multiple sequence alignment of the catalytic domains of the entire subtilase family (Siezen *et al.*, 1991; R.J.Siezen, manuscript in preparation); in this alignment the catalytic domain is defined as the sequence segment beginning at the predicted mature N-terminus and ending at the residue equivalent to the C-terminus of thermitase. Figure 3 shows only the alignment of NisP residues 29–377, numbered from the predicted N-terminus of the mature protease (Van der Meer *et al.*, 1993), with EpiP residues 114–461 and with mature thermitase and subtilisin BPN′; the sequence identity of these NisP residues is 42% (146 out of 344 residues) with EpiP, 28% (76 out of 270 residues) with thermitase and 27% (74 out of 271 residues) with subtilisin BPN′. Although the sequence identity of $<30\%$ with individual subtilases is rather low for establishing the proper overall alignment necessary for homology modelling, we stress that a much higher reliability has been obtained by using a multiple alignment of >100 subtilase sequences (Siezen *et al.*, 1991; R.J.Siezen, manuscript in preparation; data not shown).

For modelling of NisP, combined segments of thermitase and subtilisin totalling 283 residues were chosen with optimal sequence identity (30%) or similarity to NisP, as indicated in Figure 3. Compared with this combined subtilisin–thermitase framework, six residues are deleted in NisP; 72 residues are inserted in the catalytic domain, in addition to the 28 extra residues predicted at the N-terminus, at seven different positions (residues 44, 76–84, 102–105, 175–188, 218–235,

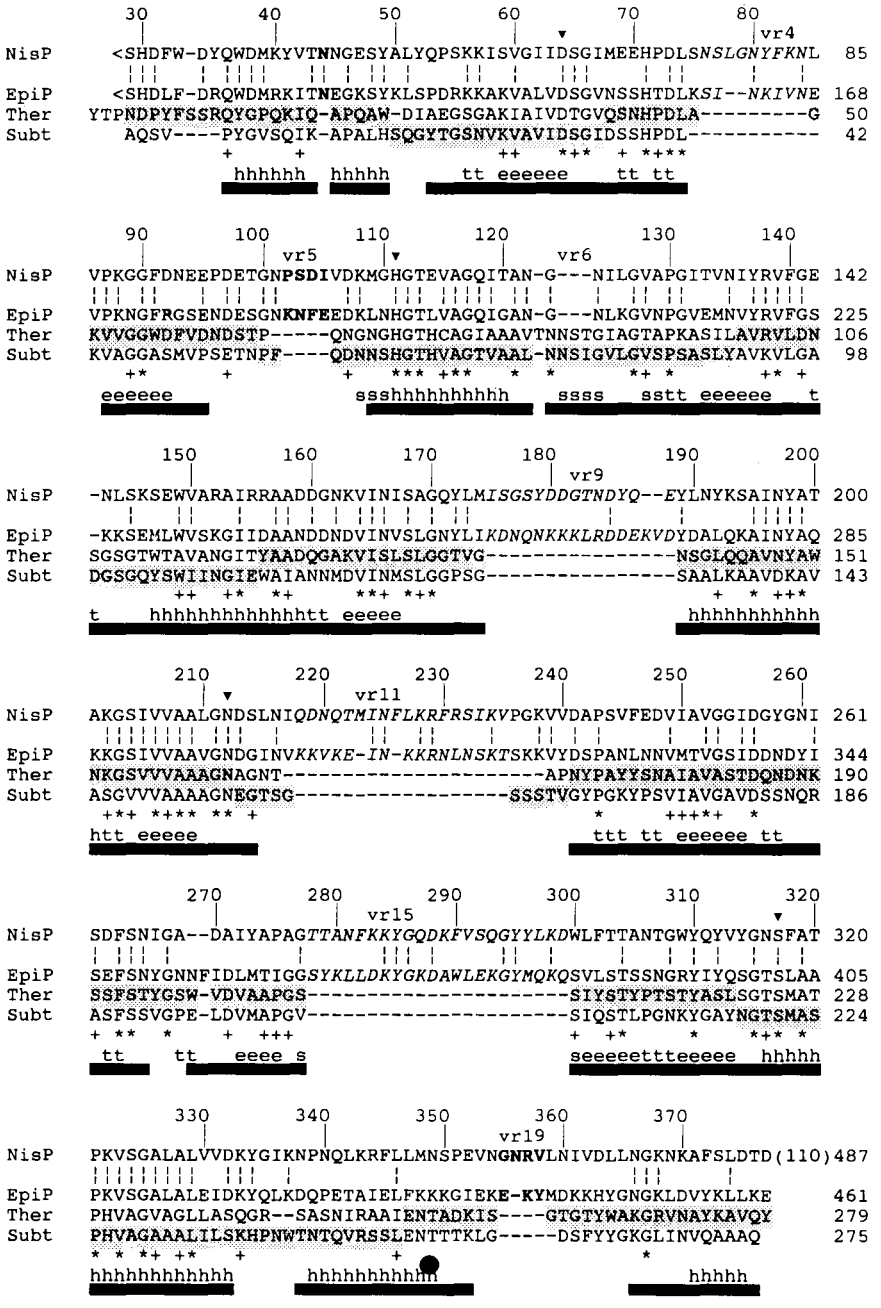

Fig. 3. Multiple sequence alignment of the proposed catalytic domain of NisP from *L.lactis* with selected members of the family of subtilisin-like serine proteases, based on the multiple alignment of >100 members of the family (Siezen *et al.*, 1991; R.J.Siezen, manuscript in preparation). EpiP, putative epidermin leader peptidase from *S.epidermidis* (Schnell *et al.*, 1992); Ther, thermitase from *Thermoactinomyces vulgaris* (Meloun *et al.*, 1985; with corrections V199W and W208S; Gros *et al.*, 1989); Subt, subtilisin BPN′ from *Bacillus amyloliquefaciens* (Wells *et al.*, 1983). Residue numbering for each enzyme is shown to the right. Numbering of thermitase, subtilisin and NisP is from the (predicted) mature N-terminus; numbering of EpiP is from the start of the pre-pro-sequence, since the mature N-terminus is unknown. Segments of thermitase and subtilisin selected as the framework for modelling of NisP are shaded. Inserted residues in NisP and EpiP relative to subtilisin and/or thermitase are shown in bold (included in modelling) or italics (not included). Structurally conserved regions in known X-ray structures are shown as solid bars, and common secondary structure elements are shown as: h, helix; e, extended β-sheet; s, bend; t, β-turn (Siezen *et al.*, 1991). Gaps are indicated by hyphens. (▼) Catalytic residues; (|) identical residues between NisP and EpiP; (*) identical residues in all four sequences; (+) identical residues in three sequences; (vr) variable regions as defined previously.

278–299 and 355–358). These insertions/deletions (indels) are all predicted to be in exposed surface loops of variable regions (vrs; numbered as before in Siezen *et al.*, 1991) between conserved secondary structural elements.

Whenever the subtilisin BPN′ residue numbering is used for reference, each number is preceded by *, e.g. the active site residues in NisP are D64(*32), H111(*64) and S317(*221).

Modelling of the catalytic domain

Molecular modelling was performed using QUANTA 3.2.3 (Molecular Simulations, Cambridge, UK) and CHARMm 22 (Brooks *et al.*, 1983) running on a Silicon Graphics 4D25TG workstation. Modelling by homology was performed essentially following standard procedures (Sutcliffe *et al.*, 1987; Blundell *et al.*, 1988; Greer, 1990, 1991; Ring and Cohen,

Table I. Ligands of the calcium binding sites in subtilisin BPN′, thermitase and proteinase K, and corresponding sites predicted in NisP and EpiP

Site	Subtilisin BPN′	Thermitase	Proteinase K	NisP	EpiP
Ca1 (strong)					
s	Q2	D5	(N5)[a]	(H30)[a]	(H)[a]
s	D41	D47	(E48)	(D73)	(D)
m	L75	V82	(R80)	(N122)	(N)
s	N77	N85	(–)	(–)	(–)
m	I79	T87	(–)	(–)	(–)
m	V81	I89	(–)	(I125)	(L)
Ca2 (medium)					
s	(S49)[a]	D57	(Q54)[a]	D92	(R)
s	(P52)	D60	(K57)	E95	E
s	(E54)	D62	(Y59)	D97	D
m	(P57)	T64	(Y61)	T99	S
s	(Q59)	Q66	(R64)	N101/D104	N/(F)
Ca3 (weak)					
m	G169	A173	A172	(S244)[a]	A
m	Y171	Y175	E174	(F246)	L
m	P172	S176	P175	(E247)	N
m	V174	A178	V176	(V249)	V
s	D197	D201	D200	(A271)	D
Ca4 (weak)					
s	(–)[a]	(–)[a]	T16	(–)[a]	(–)[a]
m	(G258)	(S260)	D260	(N354)	(K)

s, side-chain oxygen ligand; m, main-chain carbonyl ligand.
[a]Residues in brackets are predicted not to be involved in Ca ion binding; (–) indicates no topologically equivalent residue present. Subtilisin BPN′ lacks the Ca2 and Ca4 sites, thermitase the Ca4 site, and proteinase K the Ca1 and Ca2 sites. The Ca1, Ca3 and Ca4 sites are predicted to be absent in NisP. In EpiP the Ca1 and Ca4 sites are absent and the Ca2 site is predicted to be weakened or absent.

1993). The backbone conformation of NisP was modelled starting from appropriate segments of thermitase and subtilisin (Figure 3); indels of one to four residues were made by selecting fragments with the required length and best geometry, including overlap of two C_α atoms on either side of the indel, from a protein structure database. Junctions were then individually regularized by energy minimization to give reasonable geometry (50 cycles of steepest descent and 50 cycles of conjugate gradient energy minimization). The insertions of nine, 14, 18 and 22 residues (abbreviated ins9, ins14, ins18 and ins22, respectively) were not modelled since they are too large for a reliable prediction. After the substitution of side chains to obtain the desired amino acid sequence of NisP, the entire molecule was regularized in a similar fashion. Next, steric clashes were removed by the manual adjustment of side-chain torsion angles. The calcium ion Ca2 from thermitase was added, and manual adjustment of the side chains of E95, D97, D104 and R138 was carried out to optimize coordination.

Modelling of enzyme–substrate interaction

The structure of the inhibitor R45-eglin, in complex with subtilisin BPN′ (Heinz *et al.*, 1991), was chosen as a suitable template to model substrates with an Arg in the P1 position, as is the case for precursor nisin, the specific substrate. In generating these models the assumption was made that the backbone conformation of the bound substrate P8–P2′ residues is similar to that of the inhibitor eglin. The backbone conformation of R45-eglin residues 38–47 (i.e. the segment PEGSPVTR↓DL that interacts with subtilisin and thermitase in the crystal structures) was used, with the side chains substituted to resemble the appropriate P8–P2′ residues of precursor nisin, i.e. KDSGASPR↓IDhb. One of the nisin structures determined by NMR (Van de Ven *et al.*, 1991) was coupled to the C-terminal end, and the rest of the leader peptide (residues −22 to–9) was modelled as an extended chain at the N-terminal end.

Finally, the enzyme–substrate model was subjected to energy minimization after constraining the catalytic residues (D64, H111, S317), the oxyanion hole residue (N212) and the putative ligands for Ca2 (Table I). The minimized average energy of this model was negative and of a similar magnitude to that calculated for the crystal structure of subtilisin BPN′ complexed to Arg45-eglin. A Ramachandran plot of the main-chain ϕ–ψ angles in the NisP model indicates that only four non-glycine and non-proline residues are in disallowed regions, i.e. N45, A50, I217 and D241 (data not shown). Further energy minimization using molecular dynamics could improve this model, but would only be relevant if the large inserts could also be included in the model.

Protein engineering

The site-directed mutagenesis of DNA sequences encoding precursor nisin, the production and secretion by *L.lactis* strains of nisin or nisin mutants, and their purification and characterization were performed as described previously (Kuipers *et al.*, 1991, 1992; Van der Meer *et al.*, 1994). Processing of precursor nisin (and its mutants) by leader peptidase NisP was assessed by analysing the supernatant of late-log cultures by SDS–PAGE (Van der Meer *et al.*, 1994). Non-processed precursor nisin mutants were detected as 6 kDa bands (56 residues), whereas processed forms were detected as a 3.5 kDa band (34 residues) representing mature nisin.

Results and discussion

Catalytic domain model

Figure 4 shows a C_α atom trace of the model of NisP superposed on the crystal structure of thermitase. As expected

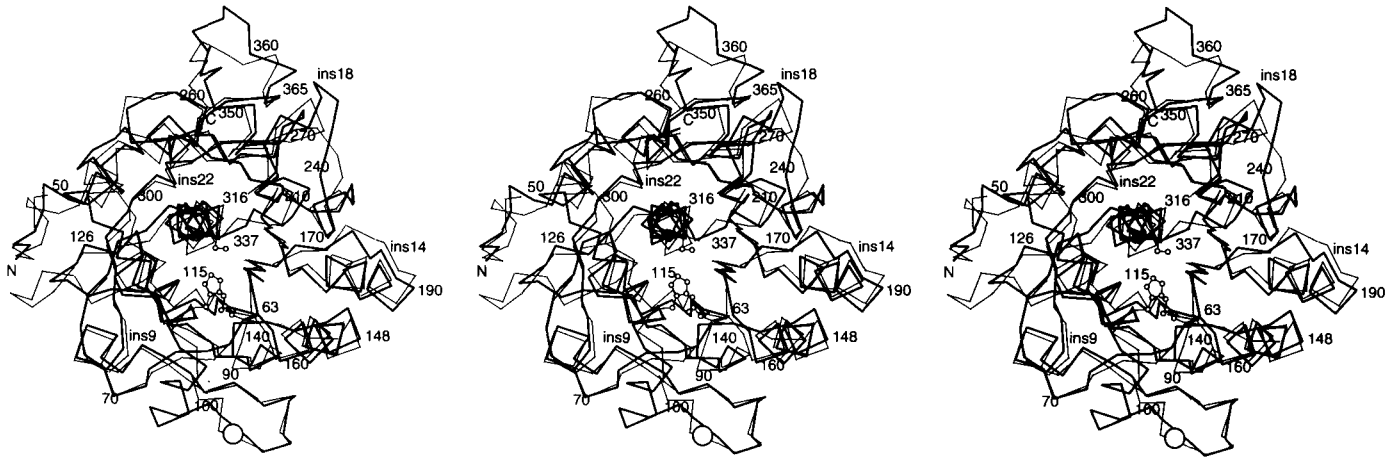

Fig. 4. Stereoview of the C_α atom trace of the NisP catalytic domain model (thick line) superimposed on the crystal structure of thermitase (thin line). Residue numbering refers to NisP. Catalytic residues D64, H111 and S317 of NisP are in small ball-and-stick representation. The large sphere indicates the Ca2 ion. N, N-terminus; C, C-terminus. Approximate positions of the four unmodelled inserts (ins9, ins14, ins18 and ins22) are shown.

(Siezen *et al.*, 1991), all of the structurally conserved regions characteristic of the subtilase family could be identified in NisP, including all of the essential α-helix and β-sheet secondary structure elements. While the framework core of the model should be quite reliable, the modelled insertions and deletions in vrs, i.e. in external loops and in connections between helices and sheet strands, are tentative. This model is oversimplified since insertions of nine (in vr4), 14 (vr9), 18 (vr11) and 22 (vr15) residues relative to thermitase/subtilisin could not be modelled (Figure 4); ins14 and ins18 are near the substrate binding region, and ins22 is near an aromatic cluster, as discussed below. These inserts could play a role in structural stabilization of the catalytic domain, in substrate binding or in interaction with other proteins. The actual conformation and the functional role of such inserted loops can only be definitively elucidated by direct X-ray structure determination. As this X-ray structure is not yet available, the present homology model provides the best current alternative for an insight into factors which determine structural stability and substrate binding in particular.

Calcium coordination sites

Four calcium ion binding sites are known from crystal structures of subtilisins, thermitase and proteinase K; these calcium ions are essential for stability and activity. From previous sequence alignments and homology modelling, it was predicted that the Ca1 (strong) and Ca3 (weak) sites are most common in members of the subtilase family, whereas the Ca2 (medium-strength) site is less common (Siezen *et al.*, 1991). The weak Ca4 site has only been found in proteinase K (Betzel *et al.*, 1988a).

Only one of the bound Ca^{2+} ions, Ca2, is predicted to be present in NisP. As shown in Table I, the coordinating ligands at the Ca2 site are very similar to those in thermitase. However, the Ca2-embracing loop in NisP is five residues longer than in thermitase (Figure 4), and an additional ligand may be provided by the negatively charged side chain of D104 in the insert. As in thermitase, the cluster of acidic residues around Ca2 is stabilized further by a conserved, buried Arg residue (Gros *et al.*, 1989), i.e. R138 in NisP.

The strong Ca1 site and weak Ca3 site, both present in thermitase and subtilisin BPN′, should not be present in NisP since the Ca1-embracing loop (in vr6) is three residues shorter and essential side-chain ligands for both Ca1 and Ca3 are missing (Table I). Proteinase K also lacks loop vr6 and hence has no Ca1 site (Betzel *et al.*, 1988a). On the other hand, novel Ca^{2+} binding sites could be present in NisP in the large inserts that were not modelled; ins14 in particular has many suitable side chains for liganding Ca^{2+} ions.

Substrate binding

Figure 5A shows a stereoview of the substrate binding region of NisP, with a P6–P2′ segment of the bound substrate precursor nisin; Figure 5B presents a schematic representation of the main features of binding. The predicted binding region in NisP can be described as a surface channel or crevice capable of accommodating at least six amino acid residues (P4–P2′) of a polypeptide substrate or pseudo-substrate (inhibitor). Both main-chain and side-chain interactions between enzyme and substrate contribute to binding. As in subtilisin and thermitase complexes, the P4–P1 backbone of the substrate is predicted to be hydrogen bonded to the NisP backbone segments 143–145 and 168–170, forming the central strand of a three-stranded antiparallel β-sheet.

In subtilisin and thermitase, the specificity appears to be largely determined by interactions of the substrate P4 and P1 residue side chains with the large, hydrophobic S4 and S1 subsites (Gron *et al.*, 1992). In NisP, the predicted subsites can be described as follows (summary in Table II): S4, a distinct pocket with the hydrophobic residues F140, V150 and A169 at one side and the more hydrophilic residues S145, S147, Q171 and Y192 (hydrogen bonded to each other) at the other side; S3, not a distinct site, since the substrate P3 residue points away from the enzyme towards the solvent; S2, a shallow pocket, bounded at the sides and bottom by F140 and active site residue H111, and at the rim by residue N143; S1, a large, elongated pocket with residue D241 at the bottom and D213 at the rim. These basic residues are expected to provide a dominant electrostatic interaction with the Arg(P1) residue of the substrate; S1′, a wide hydrophobic pocket bounded by H111 and F318 at the bottom and by M109 and V313 at the rim; S2′, a hydrophobic pocket of variable size, depending on the orientation of side chains of F264 and Y314.

This model of the NisP–precursor nisin complex suggests

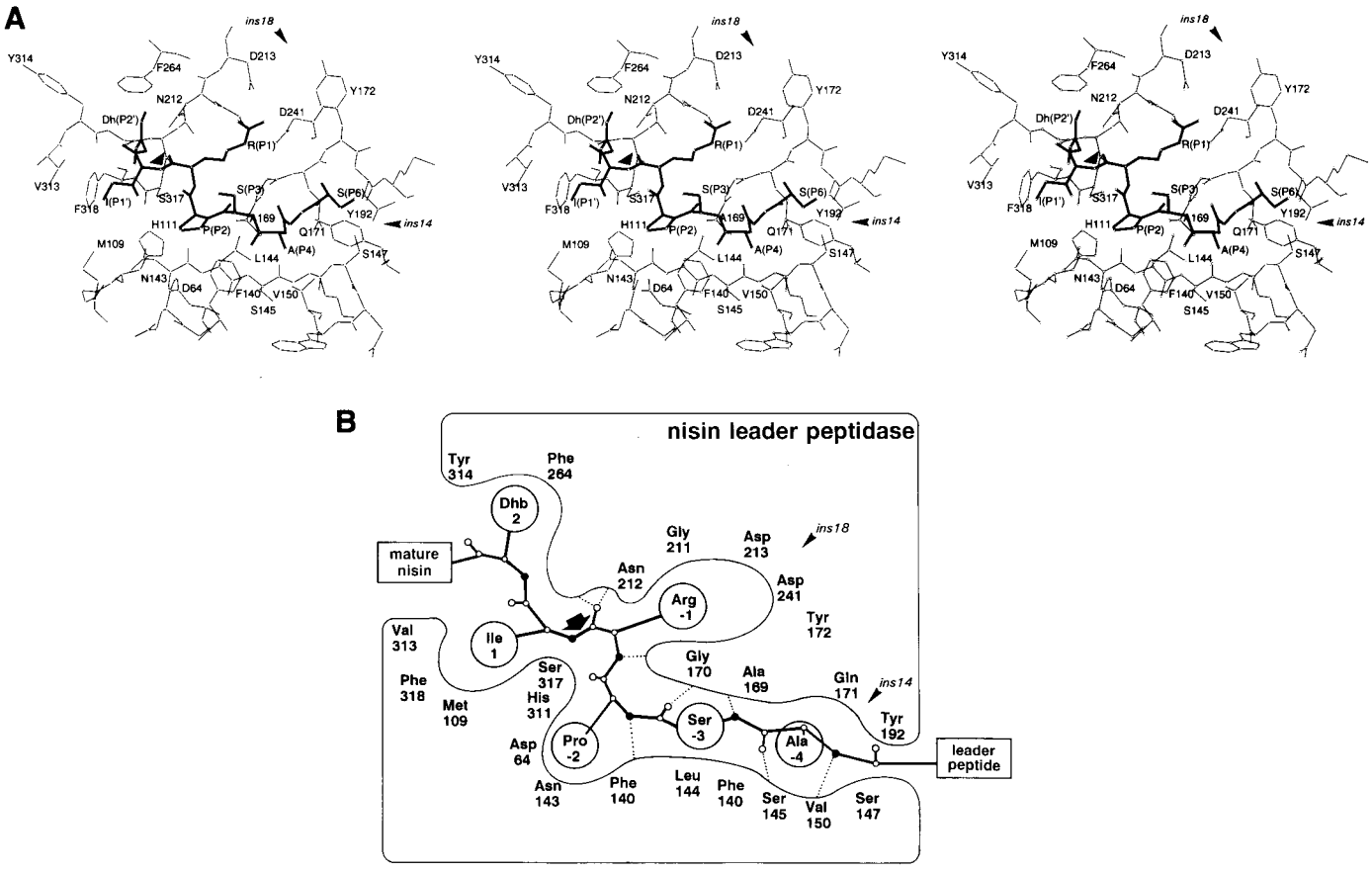

Fig. 5. (**A**) Stereoview of the modelled substrate binding region of NisP (thin lines) in complex with precursor nisin residues P6–P2′ (thick lines). (**B**) Schematic 2-D representation of the binding cleft. Substrate nomenclature P6–P2′ according to Schechter and Berger (1967). Cleavage site after the P1 residue is indicated by a large arrow, and approximate positions of unmodelled inserts by smaller arrows. Substrate backbone atoms are shown as solid spheres (nitrogen), small open spheres (carbon) and large open spheres (oxygen); potential hydrogen bonds between the substrate backbone and NisP backbone are dotted.

Table II. Possible side-chain interactions between P4–P2′ residues of precursor nisin and S4–S2′ binding sites of NisP (and equivalent residues in precursor epidermin and EpiP)

Substrate			Enzyme				
Position	Precursor nisin	Precursor epidermin	Site	NisP	EpiP	Subtilisin BPN′	
P4	A	A	S4	F140	F	L*96	side
				S145	S	G*102	side
				S147	M	Y*104	rim
				V150	V	I*107	bottom
				A169	L	L*126	side
				Q171	N	G*128	side
				Y192	L	L*135	bottom
P3	S	E	S3	L144	K	S*101	
P2	P	P	S2	H111	H	H*64	bottom
				F140	F	L*96	side
				N143	K	G*100	rim
P1	R	R	S1	G170	G	G*127	side
				Y172	Y	P*129	rim
				G211	G	G*154	side
				D213	D	E*156	rim
				D241	D	G*166	bottom
P1′	I	I	S1′	H111	H	H*64	side
				M109	L	N*62	rim
				V313	Q	Y*217	side
				F318	L	M*222	bottom
P2′	Dhb	A	S2′	F264	F	F*189	bottom
				Y314	S	N*218	side

Corresponding residues (and numbering with *) in subtilisin BPN′ are shown for reference.

that the most dominant interaction is electrostatic: that of the Arg(P1) side chain of precursor nisin with Asp residues at positions 213 and/or 241 in the S1 binding site of NisP (Figure 4 and Table II). These are precisely the two residues *156 and *166 in subtilisin that were previously changed to Glu or Asp to generate specificity for positively charged P1 residues (Wells *et al.*, 1987b). Modelling and engineering studies of pro-protein convertases have shown that specificity for Arg(P1) in this distinct subfamily of the subtilisin-like proteases is also dominated by an Asp(*166) residue at the bottom of the S1 pocket (Creemers *et al.*, 1993; Siezen *et al.*, 1994).

The S4 pocket of NisP is predicted to be smaller than that of subtilisin due to several large predominantly hydrophobic side chains (Table II). While interaction of the small residue Ala(P4) of precursor nisin in this pocket may be less important than the S1–P1 interaction, it was shown that Ala(P4) is not unfavourable for subtilisin and is still a good substrate even for the larger hydrophobic S4 pocket of subtilisin (Gron *et al.*, 1992). The Ser(P3) of precursor nisin does not appear to interact directly with NisP, although hydrogen bonding to N143 is possible. The Pro(P2) fits into a shallow S2 cleft very similar to that in subtilisin. For the segment after the cleavable bond, a stronger interaction is predicted for the Ile(P1′) residue in the hydrophobic S1′ pocket than for the Dhb(P2′) residue in the less-defined S2′ pocket.

NisP may activate itself autocatalytically by removal of its own pro-peptide, as has been demonstrated for other members of the subtilase family (Power *et al.*, 1986; Germian *et al.*, 1992; Leduc *et al.*, 1992). To investigate this possibility, the P4–P2′ residues around the putative autocatalytic cleavage site in pro-NisP, i.e. VSLR↓QP (Van der Meer *et al.*, 1993), were also modelled into the substrate binding region of NisP (results not shown). A favourable enzyme–substrate interaction is again predicted, with a dominant Arg(P1) binding and good binding of the hydrophobic Val(P4), Leu(P2) and Pro(P2′) residues in the corresponding hydrophobic clefts in NisP.

These models of substrate binding to NisP may be over-simplified, since ins14 (in vr9) and ins18 (in vr11) are near the binding region (Figures 4 and 5), and either segment may fold towards the binding cleft and contribute to substrate binding.

Taken together, however, the features of this simple model of enzyme–substrate interaction provide additional theoretical support for the observations that NisP is the natural leader peptidase for precursor nisin.

Engineering of substrate binding

The validity of the model of substrate binding can be tested by site-directed mutagenesis of enzyme or substrate, or both. The genes for precursor nisin and its leader peptidase NisP are located in a gene cluster in *L.lactis* (Kuipers *et al.*, 1993a; Van der Meer *et al.*, 1993). In substrate engineering experiments, single or multiple mutations were made in the P5–P2′ residues of precursor nisin (Table III). Substitution of the P2 residue Pro in the leader peptide to Val or Gly still produced correctly processed nisin, suggesting that these residues are equally well accommodated in the shallow S2 site of NisP. In contrast, substitution of the P4 residue Ala to Asp in precursor nisin severely reduced its processing, indicative of less effective binding to NisP. This is in agreement with our model, since the negatively charged Asp(P4) residue would not be accommodated in the hydrophobic S4 site. Furthermore, precursor nisin with a substitution of the P1 residue Arg to Gln was also not processed by NisP. Again, this supports our model of substrate binding since the Gln at P1 in precursor nisin would lead to loss of the essential electrostatic interaction with D213 and/or D241 at the bottom of the S1 site in NisP.

Replacement of the entire nisin leader peptide by that of the related lantibiotic subtilin from *Bacillus subtilis*, which ends in the P6–P1 residues SKITPQ, also inhibited cleavage by NisP (Kuipers *et al.*, 1993b). Although the P5, P4, P3 and P1 residues are simultaneously substituted in this case (Table III), the Arg(P1) to Gln substitution alone is sufficient to weaken binding to NisP, as discussed above.

Mutagenesis of the codons for the two N-terminal residues of mature nisin did not affect processing by NisP (Table III). Substitution of the P1′ residue Ile by Trp (as in subtilin) led to the production of fully processed, active I1W-nisin. Apparently, the much larger, hydrophobic side chain of Trp fits efficiently into the wide, hydrophobic S1′ site, as predicted in our model. Substitution of the codon for the P2′ residue from Thr to Ser led to the production of fully processed and active Dhb2Dha-nisin. In this case, the introduced Ser residue is found to be post-translationally modified to dehydroalanine (Dha); the Dha residue can fit into the predicted hydrophobic S2′ site of NisP since it is one methyl group smaller than the Dhb in wild-type nisin.

Aromatic surface patch

One side of the catalytic domain of NisP is predicted to be extremely rich in exposed aromatic residues (Figure 6). This

Table III. Protein engineering analysis of the interaction between precursor nisin and the leader peptidase NisP

	Precursor nisin binding region								Precursor nisin cleavage by NisP
	P6	P5	P4	P3	P2	P1 ↓	P1′	P2′	
Wild-type	Ser	Gly	Ala	Ser	Pro	Arg	Ile	Dhb	+
Mutation									
Subtilin leader		Lys	Ile	Thr		Gln			–
A(–4)D			Asp						–
P(–2)G					Gly				+
P(–2)V					Val				+
R(–1)Q						Gln			–
I1W							Trp		+
T2S								Dha	+

For experimental details see Kuipers *et al.* (1992, 1993b) and Van der Meer *et al.* (1994).

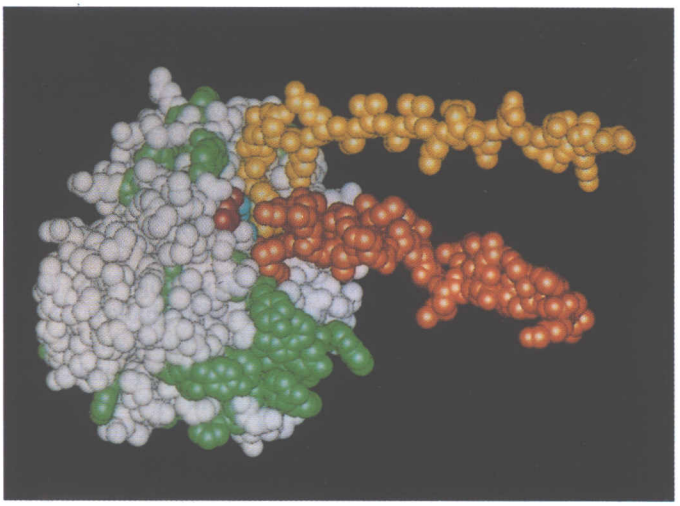

Fig. 6. Model of the catalytic domain of nisin leader peptidase NisP (white) in complex with precursor nisin (yellow, leader peptide; orange, mature nisin). Aromatic residues of NisP are shown in green. The interaction between the Arg(P1) residue (blue) of the leader with the acidic residues Asp213 and Asp241 (red) of the S1 binding pocket of NisP is shown.

unusual feature has not been observed yet in other subtilases. A total of 19 aromatics are exposed in this region, of which 14 (four Trp, four Phe and six Tyr) are clustered in a surface patch; the latter aromatics are all found in the linear sequence segments 32–41 and 258–318.

The large unmodelled insert of 22 residues, ins22(vr15), is predicted to be in the same region, i.e. between residues 277 and 300. One possibility is that ins22 merely covers the aromatic patch and buries it in the protein interior. Alternatively, ins22 may enhance the exposed aromatic patch since ins22 contains an additional two Phe and three Tyr residues. In the latter case, it is interesting to speculate what the role could be of this large aromatic surface patch on NisP. Since the enzyme is already attached to the cell membrane by the C-terminal membrane anchor (Figure 2), the aromatic patch may provide further anchoring by hydrophobic interaction with membrane lipids, or more specifically with membrane proteins such as the ABC-type translocator NisT, also encoded in the nisin operon (Engelke *et al.*, 1992; Kuipers *et al.*, 1993a). In the latter case, NisP could be part of a complex of proteins involved in biosynthesis, secretion and activation of nisin. Upon translocation, the precursor nisin would encounter NisP and be cleaved by this specific leader peptidase, leading to the release of active nisin into the medium.

A different role for the aromatic patch could be to enhance substrate binding by interacting with hydrophobic residues of the mature nisin part of the precursor prior to cleavage; the model in Figure 6 suggests that rotation around one or two bonds in the substrate would bring these two regions in close contact. This option seems less likely, however, since the specific hydrophobic interactions with the aromatic patch would need to be broken again to release mature nisin into the medium.

Other lantibiotic leader peptidases

The overall secondary and tertiary structure of the catalytic domain of the putative leader peptidase EpiP from *S.epidermidis* is predicted to be similar to that postulated for NisP (Figure 4). Of the >100 known members of the subtilase family (Siezen *et al.*, 1991; R.J.Siezen, manuscript in preparation), EpiP has the highest homology to NisP, with a 42% sequence identity and many more positions that are highly conserved with respect either to a particular residue or to charged, hydrophobic or aromatic residues (Figure 3). Minor differences in folding between EpiP and NisP are to be expected where surface loops differ by one or two residues in length, and possibly at the N-termini which are not known. No strong Ca^{2+} binding sites are predicted, but one or two weak sites may be present (Table I). The four large inserts cannot be modelled in EpiP either. While their lengths are similar to those in NisP, suggesting an essential functional role for these regions, there is very low sequence homology in these large inserts. In contrast to NisP, two of these inserts in EpiP are highly charged (19 out of 32 residues). Moreover, the aromatic surface patch is not conserved in EpiP. These differences with NisP may be related to the fact that EpiP has no extension beyond the catalytic domain, and hence no membrane anchor. EpiP apparently does not remain attached to the cell, as it has been isolated and purified from the culture filtrate of *epiP*-expressing *Staphylococcus carnosus* cells (T.Kupke, personal communication).

In the substrate for EpiP, the precursor of the lantibiotic epidermin, the only differences in the P6–P2′ residues are the substitution of Ser(P3) by Glu and Dhb(P2′) by Ala. As in NisP, the S4 pocket of EpiP is predicted to be hydrophobic, and the S1 pocket contains the same two important aspartates (Table II). Engineering of precursor epidermin has shown that, in contrast to wild-type precursor epidermin, the R(–1)Q mutant is not processed by EpiP (T.Kupke, personal communication); this is identical to the situation with R(–1)Q–precursor nisin (Table III). Interestingly, in contrast to NisP, EpiP has basic residues (Lys*100 and Lys*101) in the substrate binding region that could interact electrostatically with the Glu(P3) of pre-epidermin; this predicted P3–S3 interaction is very similar to that proposed for the cell envelope proteinase of *L.lactis* (Siezen *et al.*, 1993). Hence, binding of pre-epidermin to EpiP may be dominated by electrostatic interactions of both the Arg(P1) and Glu(P3) residues.

Conclusions

For subtilases of unknown 3-D structure, such as NisP and EpiP, homology modelling based on appropriate crystal structures (subtilisins, thermitase) provides a convenient approach to predict the overall structure of the catalytic domain, and more specifically the interactions involved in substrate binding. While the hydrophobic S1 and S4 binding pockets appear to dominate substrate binding in subtilisins (Gron *et al.*, 1992), our analysis of binding sites in NisP suggests that electrostatic interactions may alternatively contribute to binding and even dominate overall selectivity (or specificity). In particular, in the case of NisP and EpiP, one or more negative charges in the S1 pocket can lead to a high selectivity for a basic P1 residue in the substrate.

Based on these models, protein engineering can be employed not only to test the predictions, as demonstrated above for NisP, but equally importantly to design lantibiotic leader peptidases with a desired specificity or to design novel substrates or inhibitors.

Acknowledgements

We are grateful to Dr B.Renckens for expert assistance with molecular modelling, to Dr J.R.Van der Meer and H.Kosters for protein engineering, and to our colleagues for helpful discussions and editing of this manuscript. This work was supported by EC contract BIOT-CT91-0265.

References

Betzel,C., Pal,G.P. and Saenger,W. (1988a) *Eur. J. Biochem.*, **178**, 155–171.

Betzel,C., Belleman,M., Pal,G.P., Bajorath,J., Saenger,W. and Wilson,K.S. (1988b) *Proteins: Struct. Funct. Genet.*, **4**, 157–164.

Betzel,C., Teplyakov,A.V., Harutyunayn,E.H., Saenger,W. and Wilson,K.S. (1990) *Protein Engng*, **3**, 161–172.

Betzel,C., Klupsch,S., Papendorf,G., Hastrup,S., Branner,S. and Wilson,K.S. (1992) *J. Mol. Biol.*, **223**, 427–445.

Betzel,C., Singh,T.P., Visanji,M., Peters,K., Fittkau,S., Saenger,W. and Wilson,K.S. (1993) *J. Biol. Chem.*, **268**, 15854–15868.

Blundell,T., Carney,D., Gardner,S., Hayes,F., Howlin,B., Overington,J., Singh,D.A., Sibande,B.L. and Sutcliffe,M. (1988) *Eur. J. Biochem.*, **172**, 513–520.

Bode,W., Papamokos,E., Musil,D., Seemueller,U. and Fritz,H. (1986) *EMBO J.*, **4**, 813–818.

Bode,W., Papamokos,E. and Musil,D. (1987) *Eur. J. Biochem.*, **166**, 673–692.

Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaninathan,S. and Karplus,M. (1983) *J. Comput. Chem.*, **4**, 187–217.

Creemers,J.W.M., Siezen,R.J., Roebroek,A.J.M., Ayoubi,T.A.Y., Huylebroeck,D. and Van de Ven,W.J.M. (1993) *J. Biol. Chem.*, **268**, 21826–21834.

Dauter,Z., Betzel,C., Genov,N., Pipon,N. and Wilson,K.S. (1991) *Acta Crystallogr.*, **B47**, 707–730.

Delves-Broughton,J. (1990) *Food Technol.*, **44**, 100–117.

Engelke,G., Gutowski-Eckel,Z., Hammelmann,M. and Entian,K.-D. (1992) *Appl. Environ. Microbiol.*, **58**, 3730–3743.

Estell,D.A., Graycar,T.P., Miller,J.V., Powers,D.B., Burnier,J.P., Ng,P.G. and Wells,J.A. (1986) *Science*, **233**, 659–663.

Germain,D., Dumas,F., Vernet,T., Bourbonnais,Y., Thomas,D.Y. and Boileau,G. (1992) *FEBS Lett.*, **299**, 283–286.

Godette,D.W., Paech,C., Yang,S.S., Mielenz,J.R., Bystroff,C., Wilke,M.E. and Fletterick,R.J. (1992) *J. Mol. Biol.*, **228**, 580–595.

Greer,J. (1990) *Proteins*, **7**, 317–334.

Greer,J. (1991) *Methods Enzymol.*, **202**, 239–252.

Gron,H., Meldal,M. and Breddam,K. (1992) *Biochemistry*, **31**, 6011–6018.

Gros,P., Betzel,Ch., Dauter,Z., Wilson,K.S. and Hol,W.G.J. (1989) *J. Mol. Biol.*, **210**, 347–367.

Gross,E. and Morell,J.L. (1971) *J. Am. Chem. Soc.*, **93**, 4634–4635.

Heinz,D.W., Priestle,J.P., Rahuel,J., Wilson,K.S. and Grütter,M.G. (1991) *J. Mol. Biol.*, **217**, 353–371.

Hirono,S., Akagawa,H., Mitsui,Y. and Iitaka,Y. (1984) *J. Mol. Biol.*, **178**, 389–413.

Hurst,A. (1981) *Adv. Appl. Microbiol.*, **27**, 85–123.

Kuipers,O.P., Boot,H.J. and De Vos,W.M. (1991) *Nucleic Acids Res.*, **19**, 4558.

Kuipers,O.P., Rollema,H.S., Yap,W.M.G.J., Boot,H.J., Siezen,R.J. and De Vos,W.M. (1992) *J. Biol. Chem.*, **267**, 24340–24346.

Kuipers,O.P., Beerthuyzen,M.M., Siezen,R.J. and De Vos,W.M. (1993a) *Eur. J. Biochem.*, **216**, 281–291.

Kuipers,O.P., Rollema,H.S., de Vos,W.M. and Siezen,R.J. (1993b) *FEBS Lett.*, **330**, 23–27.

Lange,G., Betzel,C., Branner,S. and Wilson,K.S. (1994) *Eur. J. Biochem.*, **224**, 507–518.

Leduc,R., Molloy,S.S., Thorne,B.A. and Thomas,G. (1992) *J. Biol. Chem.*, **267**, 14304–14308.

McPhalen,C.A. and James,M.N.G. (1988) *Biochemistry*, **27**, 6582–6598.

McPhalen,C.A., Svendsen,I., Jonassen,I. and James,M.N.G. (1985) *Proc. Natl Acad. Sci. USA*, **82**, 7242–7246.

Meloun,B., Baudys,M., Kostka,V., Hausdorf,G., Frömmel,C. and Höhne,W.E. (1985) *FEBS Lett.*, **183**, 195–200.

Power,S.D., Adams,R.M. and Wells,J.A. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 3096–3100.

Rauch,P.J.G. and De Vos,W.M. (1992) *J. Bacteriol.*, **174**, 1280–1287.

Ring,C.S. and Cohen,F.E. (1993) *FASEB J.*, **7**, 783–789.

Russell,A.J. and Fersht,A.R. (1987) *Nature*, **328**, 496–500.

Russell,A.J., Thomas,P.G. and Fersht,A.R. (1987) *J. Mol. Biol.*, **193**, 803–813.

Schechter,I. and Berger,A. (1967) *Biochem. Biophys. Res. Commun.*, **27**, 157–162.

Schnell,N., Engelke,G., Augustin,J., Rosenstein,R., Ungerman,V., Götz,F. and Entian,K.-D. (1992) *Eur. J. Biochem.*, **204**, 57–68.

Siezen,R.J., de Vos,W.M., Leunissen,J.A. and Dijkstra,B.W. (1991) *Protein Engng*, **4**, 719–737.

Siezen,R.J., Bruinenberg,P.G., Vos,P., Van Alen-Boerrigter,I.J., Nijhuis,M., Alting,A.C., Exterkate,F.A. and De Vos,W.M. (1993) *Protein Engng*, **6**, 927–937.

Siezen,R.J., Creemers,J.W.M. and Van de Ven,W.J.M. (1994) *Eur. J. Biochem.*, **222**, 255–266.

Sobek,H., Hecht,H.J., Aehle,W. and Schomburg,D. (1992) *J. Mol. Biol.*, **228**, 108–117.

Sutcliffe,M.J., Haneef,I., Carney,D. and Blundell,T.L. (1987) *Protein Engng*, **1**, 377–384.

Takeuchi,Y., Noguchi,S., Satow,Y., Kojima,S., Kumagai,I., Miura,K., Nakamura,K.T. and Mitsui,Y. (1991a) *Protein Engng*, **4**, 501–508.

Takeuchi,Y., Satow,Y., Nakamura,K.T. and Mitsui,Y. (1991b) *J. Mol. Biol.*, **221**, 309–325.

Teplyakov,A.V., Kuranova,I.P., Harutyunyan,E.H., Vainshtein,B.K., Frömmel,C., Höhne,W.E. and Wilson,K.S. (1990) *J. Mol. Biol.*, **214**, 261–279.

Van de Ven,F.J.M., Van den Hooven,H.W., Konings,R.N.H. and Hilbers,C.W. (1991) *Eur. J. Biochem.*, **202**, 1181–1188.

Van der Laan,J.M., Teplyakov,A.V., Kelders,H., Kalk,K.H., Misset,O., Mulleners,L.J.S.M. and Dijkstra,B.W. (1992) *Protein Engng*, **5**, 405–411.

Van der Meer,J.R., Polman,J., Beerthuyzen,M.M., Siezen,R.J., Kuipers,O.P. and de Vos,W.M. (1993) *J. Bacteriol.*, **175**, 2578–2588.

Van der Meer,J.R., Rollema,H.S., Siezen,R.J., Beerthuyzen,M.M., Kuipers,O.P. and De Vos,W.M. (1994) *J. Biol. Chem.*, **269**, 3555–3562.

Vos,P.A.J., Simons,G., Siezen,R.J. and De Vos,W.M. (1989) *J. Biol. Chem.*, **264**, 13579–13585.

Wells,J.A., Ferrari,E., Henner,D.J., Estell,D.A. and Chen,E.Y. (1983) *Nucleic Acids Res.*, **11**, 7911–7925.

Wells,J.A., Cunningham,B.C., Graycar,T.P. and Estell,D.A. (1987a) *Proc. Natl Acad. Sci. USA*, **84**, 5167–5171.

Wells,J.A., Powers,D.B., Bott,R.R., Graycar,T.P. and Estell,D.A (1987b) *Proc. Natl Acad. Sci. USA*, **84**, 1219–1223.

Received May 24, 1994; revised October 28, 1994; accepted November 11, 1994